Statistics and Probability Letters 107 (2015) 115–121

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Inconsistent hybrid bootstrap confidence regions

Robert W. Keener[a,*], Hokeun Sun[b]

[a] Department of Statistics, University of Michigan, Ann Arbor, MI 48103, USA
[b] Department of Statistics, Pusan National University, Pusan, South Korea

ARTICLE INFO

Article history:
Received 31 August 2010
Received in revised form 17 August 2015
Accepted 17 August 2015
Available online 24 August 2015

MSC:
62F40
62F25
62E20

Keywords:
Hybrid bootstrap
Confidence intervals
Resampling
Bootstrap

ABSTRACT

In a simple signal plus noise model for Poisson data, of interest in high energy physics, coverage probabilities for hybrid bootstrap confidence intervals do not converge to the desired nominal value as information about the nuisance parameter increases.

1. Introduction

A standard approach used to find a confidence set for an unknown parameter $\theta \in \Omega$ based on data $X \sim P_\theta$ is to invert the family of likelihood ratio tests. Specifically, if $l(\cdot)$ is the log likelihood function, if $\hat{\theta}$ is the maximum likelihood estimator, if

$$\Lambda(\theta_0) = l(\hat{\theta}) - l(\theta_0)$$

is the generalized log likelihood ratio test statistic used to test $\theta = \theta_0$ versus $\theta \neq \theta_0$, and if $q(\theta)$ is the upper αth quantile for the P_θ distribution of $\Lambda(\theta)$, then

$$\{\theta : \Lambda(\theta) \leq q(\theta)\}$$

is a $1 - \alpha$ confidence region for θ.

The hybrid bootstrap, introduced to set confidence intervals in group sequential tests in Chuang and Lai (1998), is considered in a general context in Chuang and Lai (2000). The method extends the duality approach just described to interval estimation for a parameter of interest when there are nuisance parameters. Let θ and η denote the parameter of interest and the nuisance parameter, and let $\hat{\theta}$ and $\hat{\eta}$ be the maximum likelihood estimators for these parameters. If $\hat{\eta}_\theta$ maximizes the log likelihood $l(\theta, \eta)$ over η with θ fixed, then the log likelihood test statistic to test $\theta = \theta_0$ versus $\theta \neq \theta_0$ is now

$$\Lambda(\theta_0) = l(\hat{\theta}, \hat{\eta}) - l(\theta_0, \hat{\eta}_{\theta_0}).$$

* Corresponding author.
E-mail addresses: keener@umich.edu (R.W. Keener), hsun@pusan.ac.kr (H. Sun).

http://dx.doi.org/10.1016/j.spl.2015.08.014

Let $q(\theta, \eta)$ denote the upper αth quantile for $\Lambda(\theta)$ under $P_{\theta,\eta}$. The region

$$\{\theta : \Lambda(\theta) \le q(\theta, \eta),\ \forall\eta\} \tag{1}$$

has coverage probability at least $1-\alpha$. But to find it, quantiles $q(\theta, \eta)$ are needed for all θ and η, and the region may be too conservative. The ordinary bootstrap confidence region is

$$\{\theta : \Lambda(\theta) \le q(\hat\theta, \hat\eta)\}. \tag{2}$$

The only quantile necessary to compute this region is $q(\hat\theta, \hat\eta)$. This quantile can be found (if necessary) by bootstrap simulation generating data X^* from $P_{\hat\theta,\hat\eta}$. The $P_{\theta,\eta}$ coverage of this interval will be approximately $1-\alpha$ if $q(\hat\theta, \hat\eta)$ accurately estimates $q(\theta, \eta)$. In regular models with large samples this will be the case for two reasons: the maximum likelihood estimators $\hat\theta$ and $\hat\eta$ are consistent, and the null distributions for $\Lambda(\theta)$ are approximately independent of θ. In practice, the bootstrap region (2) often works well with moderate sample sizes, but with smaller samples its performance is suspect.

The hybrid bootstrap confidence region is

$$S = S(X) = \{\theta : \Lambda(\theta) \le q(\theta, \hat\eta_\theta)\}. \tag{3}$$

To compute S, quantiles $q(\theta, \hat\eta_\theta)$ are necessary for all θ. These can be found by bootstrap simulation generating data X^*_θ from $P_{\theta,\hat\eta_\theta}$ for values of θ in a reasonably fine grid. Since multiple simulations are required, the computational burden to compute the hybrid region S is greater than that for the ordinary bootstrap, especially if η is multidimensional. But with modern computing the computations are often feasible. Note that bootstrap simulations to find $q(\theta, \eta)$ for all θ and η to compute the first interval (1) would need to be done for a grid of values for θ and η, posing a greater burden than the simulations necessary for the hybrid region S in (3).

The $P_{\theta,\eta}$ coverage for the hybrid region S will be approximately $1-\alpha$ if $q(\theta, \hat\eta_\theta)$ accurately estimates $q(\theta, \eta)$. As with the bootstrap region, this should be the case in large samples but may be suspect with small samples. But there are several interesting examples in which the data provide substantial information about the nuisance parameter η, but limited information about the parameter of interest θ. In these cases, the hybrid region may perform much better than the ordinary bootstrap region.

The next section considers a simple signal plus noise model for Poisson data, of interest in high energy physics. In this example coverage probabilities for the hybrid region S do not converge to the desired nominal value as information about the nuisance parameter increases. Inconsistency arises because test statistics for θ are discrete when η is known.

When the nuisance parameter η is known, exact coverage can be achieved in a standard fashion by introducing extraneous randomization. Although this gives smaller intervals and increases the sensitivity of associated tests, randomization feels artificial and this approach may not have much appeal. The situation is similar when η is unknown. If desired, estimation error for $\hat\eta$ can used to construct variables that are approximately uniformly distributed on $(0, 1)$; and these variables can be used to construct consistent confidence intervals analogous to the exact intervals based on external randomization when the η is known. But this approach uses $\hat\eta$ a rather discontinuous fashion and feels artificial. Unfortunately, intervals that use the data in a more natural fashion will be inconsistent. Similar difficulties will almost certainly arise in other problems that are discrete when the nuisance parameter is known.

2. Poisson example

Researchers in high energy physics are at times interested in estimating a rate $\theta \ge 0$ from a Poisson measurement X with mean $\theta + \eta$. Here η represents a background rate, often considered as known from prior or "off-line" experiments. Also, in many cases $\theta = 0$ is a definite possibility, corresponding to the absence of the particle or phenomena the experiment is trying to detect. This problem is a bit nonstandard since EX is known to be at least η, and there has been some discussion in the physics literature about the proper way to set a confidence interval for θ. The "unified method" of Feldman and Cousins (1998) amounts to inverting the family of likelihood ratio tests, and has seen wide interest in physics since its appearance. Related alternatives are discussed in Mandelkern (2002), Roe and Woodroofe (1999), and Roe and Woodroofe (2000).

In practice, the assumption that the background rate η is known is too optimistic. More realistically, information about η may come from count data Y modeled as Poisson with mean $\gamma\eta$. Here the scale factor γ, represents the ratio of the observation times for Y and X. With large γ there is considerable information about the background η, exactly the setting in which the hybrid bootstrap approach seems most promising.

Sen et al. (2009) have investigated the performance of the hybrid bootstrap confidence interval S in this example, and in their numerical work it seems to perform well. Feldman (2000) also considers this problem, although it is not clear if his suggestions are exactly the same as the hybrid bootstrap.

Despite the positive accounts of the hybrid bootstrap interval's performance, in this example S is not consistent: as $\gamma \to \infty$, $P_{\theta,\eta}(\theta \in S) \not\to 1-\alpha$. The problem has to do with discreteness. Note that if the background η were known, exact coverage based on data X would be impossible without external randomization.

For notation, let

$$l_0(\theta, \eta) = X\log(\theta + \eta) - \theta - \eta - \log(X!),$$